

Transient Analysis of D(t)/M(t)/1 Queuing System with Applications to Computing Airport Delays

by

Shubham Gupta

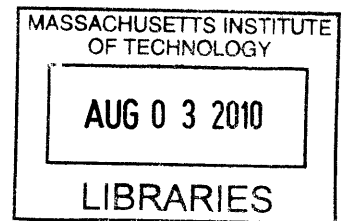
B.Tech, Computer Science and Engineering (2007),
Indian Institute of Technology (IIT), Kanpur

Submitted to the Sloan School of Management
in partial fulfillment of the requirements for the degree of
Master of Science in Operations Research
at the

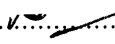
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2010


©Massachusetts Institute of Technology 2010. All rights reserved.




ARCHIVES

Author.....

Shubham Gupta
Operations Research Center, MIT.

Certified by.....

Amedeo Odoni
Professor of Aeronautics and Astronautics,
Professor of Civil and Environmental Engineering,
Thesis Supervisor.

Accepted by.....

Dimitris J. Bertsimas
Boeing Professor of Operations Research, Sloan School of Management,
Co-director, Operations Research Center.

Transient Analysis of $D(t)/M(t)/1$ Queuing System with Applications to Computing Airport Delays

by

Shubham Gupta

B.Tech, Computer Science and Engineering (2007),
Indian Institute of Technology (IIT), Kanpur

Submitted to the Sloan School of Management
in partial fulfillment of the requirements for the degree of
Master of Science in Operations Research.

Abstract

This thesis is motivated by the desire to estimate air traffic delays at airports under a range of assumptions about the predictability of (a) inter-arrival times of demands (arrivals and departures) and (b) service times of aircraft movements (landings and takeoffs). It consists of two main parts.

In the first, a transient analysis of a $D(t)/M(t)/1$ queuing system is presented. The reason for focusing on such a system is that it may be useful in evaluating some of the benefits of a future Air Traffic Management (ATM) system, such as the Next Generation Air Transportation System (NGATS or NextGen) currently being developed in the United States. One of the main features of these future ATM systems will be high predictability and regularity of the inter-arrival times of airport demands, i.e., a nearly deterministic demand process. This will be achieved through significant reductions in aircraft trajectory uncertainty, with the expectation that airport delays will also decrease substantially as a result. We develop a novel, computationally-efficient numerical approach for solving $D(t)/M(t)/1$ queuing systems with general, dynamic demand and service rates. We also discuss the complexity of the approach and some characteristics of the resulting solutions.

In the second part of the thesis, we use a set of models of dynamic queuing systems, in addition to our $D(t)/M(t)/1$ model to explore the range of values that airport delays may assume under different sets of assumptions about the level of uncertainty associated with demand inter-arrival times and with service times. We thus compute airport delays under different queuing systems in a dynamic setting (where demand and service rates are time-varying) to capture the entire range of uncertainties expected during the deployment of various future ATM system technologies. The specific additional models we consider are: a deterministic $D(t)/D(t)/1$ model in which it is assumed that airport demands for landings and takeoffs occur at exactly as scheduled; and a $M(t)/E_k(t)/1$ model which, because of the “richness” of the family of Erlang distributions, E_k , can be used to approximate most $M(t)/G(t)/1$ models

that may arise in airport applications. It can be seen that these models, when used together, provide bounds on estimated airport delays, with the $D(t)/D(t)/1$ model most likely to offer a lower bound and the $M(t)/M(t)/1$ model (i.e., the special case of $M(t)/E_k(t)/1$ with $k = 1$), an upper bound. We show through a set of examples based on a few of the busiest airports in the United States that: the closeness of the delay estimates provided by the different models depend on the level of congestion at an airport and the relative shapes of the dynamic profiles of capacity and demand at the airport; the difference (on a “percentage” basis) between the estimates provided by the deterministic model and the stochastic ones is largest for uncongested airports and decreases as the level of congestion increases; $D(t)/M(t)/1$ and $M(t)/D(t)/1$ produce estimates of the same order of magnitude, and reflect delays in the presence of “moderate” uncertainty at an airport; and delays under a $D(t)/M(t)/1$ queuing system are always higher than under a $M(t)/D(t)/1$ system.

Thesis Supervisor: Amedeo Odoni

Title: Professor of Aeronautics and Astronautics, Professor of Civil and Environmental Engineering

Acknowledgments

I would like to express my deepest gratitude to my supervisor, Professor Amedeo Odoni, who has guided me and inspired me through my research. I shall always cherish his mentorship over the course of this thesis. I feel fortunate to join the illustrious list of his students.

I express my gratitude to Professor Dimitris Bertsimas for his support and care. I thank David Goldberg for some useful discussions on the work presented in this thesis. I would like to thank my parents for always providing me with the best of opportunities. All this would be futile without their love and care.

To my parents

Contents

1	Introduction	4
1.1	Background	5
1.2	Related Literature	6
1.3	Thesis Overview	7
2	Analysis of $D(t)/M(t)/1$ Queuing System	9
2.1	Time-varying demand rate and constant service rate	10
2.2	Time-varying demand and service rate	12
2.3	Some Performance Characteristics	14
2.3.1	Complexity	15
2.3.2	Comparison of Steady State Statistics with $M/E_k/1$ System .	17
2.3.3	Comparison of Transient Delays with the $M(t)/D(t)/1$ System	19
2.4	Summary	22
3	Comparing Airport Delay Estimates from Various Queuing Systems	23
3.1	Other Queuing Systems	23
3.1.1	The $M(t)/E_k(t)/1$ System	23
3.1.2	Two Versions of a $D(t)/D(t)/1$ System	25
3.2	Delays Under Different Queuing Systems	26
3.3	Additional Comparisons	35
3.3.1	Effect of Congestion	35
3.3.2	Some Anomalous Cases	39
3.3.3	Regression as a Function of Level of Congestion.	41
3.4	Summary	42
4	Conclusion	43

List of Figures

1.1	The four queuing systems.	6
2.1	Explanation of effective service rate (μ^{eff}).	14
2.2	Comparison of average delay (\bar{W}) in steady-state with load factor.	18
2.3	Comparison of average number of customer in queue (\bar{L}) in steady-state with load factor.	19
3.1	Capacity Scenarios at ATL.	30
3.2	Capacity Scenarios at BOS.	31
3.3	Capacity Scenarios at LGA.	31
3.4	Capacity Scenarios at MIA.	32
3.5	Capacity Scenarios at ORD.	32
3.6	Capacity Scenarios at SFO.	33
3.7	Regression Between D/M/1 and M/D/1 Queuing Systems.	34
3.8	Regression between Micro and Macro.	36
3.9	Regression between DELAYS and Micro.	36
3.10	Regression between DELAYS and Macro.	37
3.11	Regression between DELAYS-EFF and Micro.	37
3.12	Regression between DELAYS-EFF and Macro.	38
3.13	One of the outlier points (LGA).	41

List of Tables

2.1	Comparison of average waiting time (\bar{W}) in steady-state under different load factors (ρ).	18
2.2	Comparison of expected number in the queue (\bar{L}) in steady-state under different load factors (ρ).	18
2.3	Effect of level of congestion.	20
2.4	Effect of demand variability over time (Test Case 1: $\rho = 0.5$).	21
2.5	Effect of demand variability over time (Test Case 2: $\rho = 0.8$).	21
3.1	Comparison of aggregate delays for a 17-hour period (in min) under different queuing systems - ATL.	28
3.2	Comparison of aggregate delays for a 17-hour period (in min) under different queuing systems - BOS.	29
3.3	Comparison of aggregate delays for a 17-hour period (in min) under different queuing systems - MIA.	29
3.4	Comparison of aggregate delays for a 17-hour period (in min) under different queuing systems - ORD.	29
3.5	Comparison of aggregate delays for a 17-hour period (in min) under different queuing systems - SFO.	30
3.6	Comparison of Delays (averaged over all capacity scenarios) computed for the D/M/1 and M/D/1 Systems.	30
3.7	Regression analysis as a function of level of congestion.	42

Chapter 1

Introduction

Currently, in the United States, there are significant resources being deployed in the development of a future Air Traffic Management (ATM) system called the Next Generation Air Transportation System (NGATS or NextGen). The expected key benefits of this new ATM system are improvements in various aspects of aviation encompassing ground/terminal operations, technological advancements in aircraft monitoring, efficient coordination of various ATM tools and improved tracking of actual flight paths. In particular, a major proposal of NGATS is improvements in 4DT trajectory uncertainty (4DT capability is defined as the ability to precisely fly an assigned 3D trajectory while meeting specified timing constraints on arrival at waypoints [5]). This would lead to enhanced predictability and control over the trajectory of an aircraft with the resulting decrease in the delays in the system. Thus, to quantify these benefits, it might be desirable to estimate air traffic delays at airports under a range of assumptions about the predictability of (a) inter-arrival times of demands (arrivals and departures) and (b) service times of aircraft movements (landings and takeoffs). This thesis addresses exactly this issue. It consists of two main parts:

In the first part, we present a transient analysis of a dynamically evolving $D(t)/M(t)/1$ system where the demand and service may be strongly time-varying. One of the main features of this future ATM system will be high predictability and regularity of the inter-arrival times of airport demands, i.e., a nearly deterministic demand process. Thus, the results from the $D(t)/M(t)/1$ system will help quantify the resulting benefits. Our analysis is based on computing the probabilities of the state of the system at various arrival time-points and using it to compute the

expected delays of each aircraft.

In the second part of the thesis, we use a set of models of dynamic queuing systems, in addition to our $D(t)/M(t)/1$ model to explore the range of values that airport delays may assume under different sets of assumptions about the level of uncertainty associated with demand inter-arrival times and with service times. The specific additional models we consider are: a deterministic $D(t)/D(t)/1$ model in which it is assumed that airport demands for landings and takeoffs occur at exactly as scheduled; and a $M(t)/E_k(t)/1$ model which, because of the “richness” of the family of Erlang distributions, E_k , can be used to approximate most $M(t)/G(t)/1$ models that may arise in airport applications. It can be seen that these models, when used together, provide bounds on estimated airport delays, with the $D(t)/D(t)/1$ model most likely to offer a lower bound and the $M(t)/M(t)/1$ model (i.e., the special case of $M(t)/E_k(t)/1$ with $k = 1$) an upper bound. We thus compute airport delays under different queuing systems in a dynamic setting (where demand and service rates are time-varying) to capture the entire range of uncertainties expected during the deployment of various future ATM system technologies.

1.1 Background

An airport can be modeled as a queuing system where the aircraft can be interpreted as the arriving customers requesting service to land/take-off from the runways. At an airport, the demand is in general a strongly varying function of time as there are very few aircraft arriving/departing during early morning and late night hours. This time-varying demand profile is an essential characteristic at any airport. Moreover, capacity during some time-periods may also change, as a result of weather disturbances. This, in turn, makes the “steady-state” queuing results a bad approximation to the actual delays. This is because classical queuing theory deals with computing delays in the long-run when the queuing system is allowed to operate under constant demand and service rates for a very long period of time. In contrast, any appropriate analysis of an airport modeled as a queuing system has to be carried out during its transient phase and not in “steady-state”. Thus, to carry out an analysis of the possible delays at an airport under different queuing systems, it is essential to have a portfolio of tools that aid in their exact or reasonably approximate transient analysis. One of the contributions of this thesis is in assembling such a toolset. Next, we review the various possible models of a queuing system and highlight the ones

we focus on in this thesis.

The most general setting for a dynamic queuing system is $G(t)/G(t)/1$, where the two $G(t)$'s represent the arbitrary distribution of the inter-arrival time and the service time respectively. In this thesis, we focus on the following four systems - $D(t)/D(t)/1$, $M(t)/D(t)/1$, $D(t)/M(t)/1$ and $M(t)/M(t)/1$. The $D(t)/D(t)/1$ system represents the completely deterministic setting where there is no uncertainty in both the arrival and service process. $M(t)/M(t)/1$ represents the system at the other end of the spectrum where the inter-arrival times and service times are Poisson (completely random). The other two systems ($D(t)/M(t)/1$ and $M(t)/D(t)/1$) capture the “intermediate precision” cases, i.e., we study the delays resulting from the assumptions that the arrival process ($D(t)/M(t)/1$) and the service process ($M(t)/D(t)/1$) are deterministic respectively. Figure 2.1 depicts the hierarchy of the four models that we study.

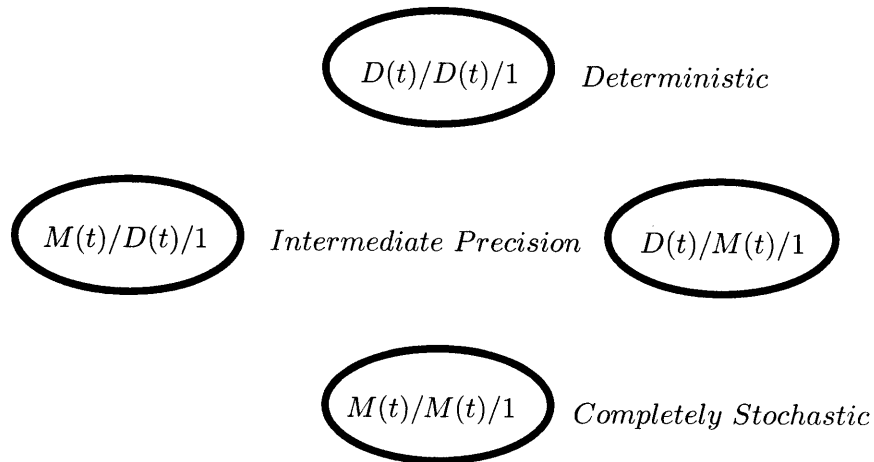


Figure 1.1: The four queuing systems.

1.2 Related Literature

Much of the theoretical analysis presented in this thesis falls within the scope of queuing theory. Although, there is a significant body of research dealing with steady-

state results (when the system is allowed to operate for a long time so as to reach equilibrium), the literature related to the transient analysis of systems with strongly time-varying demand and service has been rather sparse. Odoni, Roth [1] and Green, Kolesar [7] point out the inadequacies of steady state results for approximating the behaviour of transient and dynamic queuing systems. Most attempts to explicitly incorporate dynamic and transient behaviour model the service and arrival processes as phase-type and attempt to solve the resulting forward Kolmogorov equations. In applications of this general type of approach to the dynamic analysis of airport queues, Peterson, Bertsimas and Odoni [9] develop a model based on a Markov/semi-Markov treatment of changes in the weather (to model the uncertainty in service), together with a treatment of the arrival stream as time-varying but deterministic. Moreover, they compute moments of queue length and waiting time via a recursive algorithm. Next, we review the literature on the analysis of $D(t)/M(t)/1$ systems. Pack [8] obtains closed-form expressions for the distributions and first two moments of three random variables (namely, time until the n^{th} departure, the number of departures in an interval of length t and the time between the $(n - 1)^{\text{th}}$ and n^{th} departure) associated with the nonequilibrium output of a $D/M/1$ queue. Moreover, he also derives expressions for the equilibrium distributions and moments of these random variables. Birger [4] presents the transient analysis of a static $D/M/1$ system. To the best of our knowledge, we are not aware of any model (let alone a computationally tractable one) to solve the dynamic $D(t)/M(t)/1$ system. Our principal theoretical contribution is the development of a transient analysis of $D(t)/M(t)/1$ system which relies on computing the probabilities of the queue length at various aircraft arrival epochs. Moreover, we propose a recursive way to compute the state probabilities to make the procedure tractable. Finally, we propose the concept of an *effective service rate* which is helpful in computing delays when the service rate varies strongly with time.

1.3 Thesis Overview

This thesis is structured as follows: Chapter 2 presents the transient analysis of a $D(t)/M(t)/1$ system under both a constant and a time-varying service rate. Moreover, it presents the complexity analysis (both time and space) of the proposed approach and a comparison of transient and steady-state delays with those of $M/D/1$ systems. Chapter 3 presents a brief description of the queuing systems used for the

estimation of delays at an airport in addition to $D(t)/M(t)/1$. It then reports extensive computational results of the delays obtained under the four queuing systems on a set of US airports. Chapter 4 summarizes conclusions.

Chapter 2

Analysis of $D(t)/M(t)/1$ Queuing System

In this chapter, we present our numerical method for the transient analysis of a $D(t)/M(t)/1$ system under both constant and time-varying service rates. In addition, we analyze the complexity (both time and space) of our proposed approach; compare the transient and steady state delays with those of the $M(t)/D(t)/1$ system; and finally, conclude with a study of the effect of congestion and time-variation of demand on the system delays.

Notation.

Here, we summarize the notation used throughout this chapter:

- \mathcal{T} : total number of time-periods,
- Δ : duration of each time-period,
- α_i : total number of aircraft arrivals during time-period $i \in \mathcal{T}$,
- μ_i : service rate during time period $i \in \mathcal{T}$,
- D_k : delay experienced by the k^{th} aircraft,
- X_k : service time of the k^{th} aircraft,
- Z_k : number of aircraft in front at the time of arrival of the k^{th} aircraft,
- t_k : time-period in which the k^{th} aircraft arrives,

- c_k : inter-arrival time between k^{th} and $(k + 1)^{th}$ aircraft,
- $C_k = \sum_{i=1}^{k-1} c_i$: time of arrival of the k^{th} aircraft,
- A_c^μ : number of arrivals in an interval of length c in a Poisson process of rate μ ,
- $N = \sum_{i=1}^T \alpha_i$: total number of aircraft arrivals,
- $\mathcal{D} = \sum_{k=1}^N D_k$: total delay experienced across all aircraft.

2.1 Time-varying demand rate and constant service rate

Here, we analyze the $D(t)/M(t)/1$ system under the assumption that the demand is time-varying but the service rate is constant (denoted by μ).

Given Z_k , the total time that the k^{th} aircraft will spend in the system is the sum of Z_k i.i.d exponentials plus its own service time. Hence, the delay experienced by the k^{th} aircraft is Erlang of order Z_k (i.e., $D_k|Z_k$: Erlang of order Z_k).

$$E[D_k|Z_k] = Z_k/\mu \quad (2.1)$$

$$E[D_k] = E[E[D_k|Z_k]] \quad (2.2)$$

$$= E[Z_k]/\mu \quad (2.3)$$

$$Z_k = \begin{cases} [0, Z_{k-1}], & \text{if } X < c_{k-1}, \\ Z_{k-1} + 1, & \text{if } X \geq c_{k-1}. \end{cases} \quad (2.4)$$

where X : service time of a random aircraft.

We model this system as a discrete-time Markov Chain with time-varying transition probabilities p_{ij}^k .

$$p_{ij}^k = \begin{cases} P(Z_k = j|Z_{k-1} = i), & \text{if } j \leq i+1, \\ 0, & \text{otherwise.} \end{cases} \quad (2.5)$$

For the case when $j \leq i + 1$, p_{ij}^k can be computed as follows:

$$p_{ij}^k = P(Z_k = j | Z_{k-1} = i) \quad (2.6)$$

$$= P(A_{c_{k-1}}^\mu = i + 1 - j | A_{c_{k-1}}^\mu \leq i + 1) \quad (2.7)$$

$$= \frac{P(A_{c_{k-1}}^\mu = i + 1 - j \wedge A_{c_{k-1}}^\mu \leq i + 1)}{P(A_{c_{k-1}}^\mu \leq i + 1)} \quad (2.8)$$

$$= \frac{P(A_{c_{k-1}}^\mu = i + 1 - j)}{P(A_{c_{k-1}}^\mu \leq i + 1)} \quad (2.9)$$

$$= \frac{e^{-\mu c_{k-1}} (\mu c_{k-1})^{(i+1-j)} / (i+1-j)!}{\sum_{z=0}^{i+1} e^{-\mu c_{k-1}} (\mu c_{k-1})^z / z!} \quad (2.10)$$

$$= \frac{(\mu c_{k-1})^{(i+1-j)} / (i+1-j)!}{\sum_{z=0}^{i+1} (\mu c_{k-1})^z / z!} \quad (2.11)$$

It is important to note that the resulting state space might be large. Hence, to ease the computations involved (with the factorials), we give a recursive definition of the probabilities:

$$p_{ij}^k = \frac{i + 2 - j}{\mu c_{k-1}} p_{i,j-1}^k \quad (2.12)$$

Let $q_{k,j} = P(Z_k = j)$ denote the probability that the k^{th} aircraft sees j aircraft in the queue when it arrives. Based on the probabilities $q_{k,j}$, the delay that it experiences can be computed as follows:

$$E[D_k] = \sum_{j=1}^N \frac{j \cdot q_{k,j}}{\mu} \quad (2.13)$$

Next, we give a recursive definition for the computation of the probabilities $q_{k,j}$:

$$q_{k,j} = P(Z_k = j) \quad (2.14)$$

$$= \sum_{i=0}^N P(Z_k = j | Z_{k-1} = i) \cdot P(Z_{k-1} = i) \quad (2.15)$$

$$= \sum_{i=0}^N p_{ij}^k \cdot q_{k-1,i} \quad (2.16)$$

Given the above recursive relation between $q_{k,j}$ and $q_{k-1,j}$, and the initial condition that $q_{1,0} = 1$, all $q_{k,j}$'s and hence, all the delays can be computed.

2.2 Time-varying demand and service rate

In this section, we add the complication of time-varying service rates. In this case, the service times of all the aircraft are no longer i.i.d, and they depend on the time-interval in which an aircraft enters the queue because of time-varying service rates. Since the arrival process is still deterministic, we know the exact time of arrival of each aircraft, and hence we know the service rate with which every aircraft will get processed if it arrived in an empty queue.

We introduce some added notation to aid our subsequent analysis:

Added Notation.

- $\tilde{\mu}_k$: service rate at the instant of arrival of the k^{th} aircraft.

Note that μ_i still denotes the service rate during time-period i , but we need $\tilde{\mu}_k$ to take into account the arrival times of various aircraft. Moreover, given the exact time of arrival of aircraft k , $\tilde{\mu}_k$ can be mapped to a corresponding μ_i by mapping the arrival time to the appropriate time-period.

An exact computation of p_{ij}^k is as follows:

$$p_{ij}^k = P(Z_k = j | Z_{k-1} = i) \quad (2.17)$$

$$= P(A_{c_{k-1}}^{\tilde{\mu}_{k-1}} = i + 1 - j | A_{c_{k-1}}^{\tilde{\mu}_{k-1}} \leq i + 1) \quad (2.18)$$

$$= \frac{P(A_{c_{k-1}}^{\tilde{\mu}_{k-1}} = i + 1 - j \wedge A_{c_{k-1}}^{\tilde{\mu}_{k-1}} \leq i + 1)}{P(A_{c_{k-1}}^{\tilde{\mu}_{k-1}} \leq i + 1)} \quad (2.19)$$

$$= \frac{P(A_{c_{k-1}}^{\tilde{\mu}_{k-1}} = i + 1 - j)}{P(A_{c_{k-1}}^{\tilde{\mu}_{k-1}} \leq i + 1)} \quad (2.20)$$

$$= \frac{e^{-\tilde{\mu}_{k-1}c_{k-1}}(\tilde{\mu}_{k-1}c_{k-1})^{(i+1-j)}/(i+1-j)!}{\sum_{z=0}^{i+1} e^{-\tilde{\mu}_{k-1}c_{k-1}}(\tilde{\mu}_{k-1}c_{k-1})^z/z!} \quad (2.21)$$

$$= \frac{(\tilde{\mu}_{k-1}c_{k-1})^{(i+1-j)}/(i+1-j)!}{\sum_{z=0}^{i+1} (\tilde{\mu}_{k-1}c_{k-1})^z/z!} \quad (2.22)$$

Again, to ease the computation of the various probabilities, we give a recursive definition for p_{ij}^k as follows:

$$p_{ij}^k = \frac{i+2-j}{\tilde{\mu}_{k-1}c_{k-1}} p_{i,j-1}^k \quad (2.23)$$

Let $q_{k,j} = P(Z_k = j)$ denote the probability that the k^{th} aircraft sees j aircraft in the queue when it arrives.

$$q_{k,j} = P(Z_k = j) \quad (2.24)$$

$$= \sum_{i=0}^N P(Z_k = j | Z_{k-1} = i) \cdot P(Z_{k-1} = i) \quad (2.25)$$

$$= \sum_{i=0}^N p_{ij}^k \cdot q_{k-1,i} \quad (2.26)$$

A major difference in the analysis of this case is that, in computing the expected delays for each aircraft, the time-varying service rates have to be taken into account, that is, the queue in front of each arriving aircraft may be dissipated with different service rates. Hence, to take this into account, we introduce the notion of an *effective service rate* (μ^{eff}) to process the queue.

Effective Service Rate.

Here, we give the formal definition of effective service rate. Let $\mu_{k,j}^{eff}$ denote the effective service rate of servicing j aircraft in the queue at the time of arrival of the k^{th} aircraft. The effective service rate takes into account the different service rates. Let n_k^j denote the number of time-periods that it takes to service all the j aircraft starting at the time of arrival of the k^{th} aircraft.

$$n_k^j = \operatorname{argmin}_m \left\{ m \mid (t_k - \frac{C_k}{\Delta})\mu_{t_k} + \sum_{i=t_k+1}^m \mu_i \geq j \right\} \quad (2.27)$$

Let N_k^j denote the number of aircraft that are serviced in all but the last time-period among the n_k^j time-periods, i.e.,

$$N_k^j = \left((t_k - \frac{C_k}{\Delta})\mu_{t_k} + \sum_{i=t_k+1}^{n_k^j-1} \mu_i \right) \quad (2.28)$$

Then, $\mu_{k,j}^{eff}$ can be computed as follows:

$$\frac{j}{\mu_{k,j}^{eff}} = (t_k - \frac{C_k}{\Delta}) + (n_k^j - t_k - 1) + (\frac{j - N_k^j}{\mu_{n_k^j}}) \quad (2.29)$$

$$i.e., \mu_{k,j}^{eff} = \frac{j}{(t_k - \frac{C_k}{\Delta}) + (n_k^j - t_k - 1) + (\frac{j - N_k^j}{\mu_{n_k^j}})} \quad (2.30)$$

Figure 2.1 gives an example to elaborate on the concept of effective service rate. There are three time-periods ($\mathcal{T} = 3$) and flight f_1 lands at the middle of the first time-period. Moreover, the service rates during the entire time horizon are $\mu_1 = 20, \mu_2 = 40$ and $\mu_3 = 30$. Suppose, the queue consists of 65 aircraft at the time of arrival of flight f_1 . Then, the first 10 aircraft will be serviced with a rate of μ_1 , next 40 by μ_2 , and the last 15 by μ_3 . For this example, $n_k^j = 3, N_k^j = 50, t_k = 1$ and $C_1 = \Delta/2$. Hence, $\mu_{1,65}^{eff}$ can be computed as follows:

$$\frac{65}{\mu_{1,65}^{eff}} = \frac{1}{2} + 1 + \frac{1}{2} \quad (2.31)$$

$$i.e., \mu_{1,65}^{eff} = \frac{65}{2} = 37.5 \quad (2.32)$$

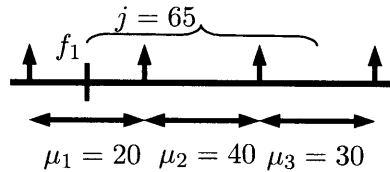


Figure 2.1: Explanation of effective service rate (μ^{eff}).

2.3 Some Performance Characteristics

In this section, we characterize the time and space complexity of our proposed numerical approach, and present a comparison of steady-state and transient delays of this queuing system for different load factors with the M(t)/D(t)/1 system.

2.3.1 Complexity

Time Complexity.

Here, we characterize the time complexity of the analysis of a $D(t)/M(t)/1$ system presented in the previous section. It should be evident that the main effort in the analysis of this system lies in the probability computations (the effort within one such computation and the number of such computations). By providing a recursive way to compute successive probabilities, we are able to reduce drastically the effort needed for each such individual probability computation. In fact, this is $O(1)$ because each probability is computed as a constant multiplied by the immediately previous computed probability. Next, we analyze the number of such probability computations involved. We introduce the following notation:

Notation.

- $t_{i,j,k}$: number of operations for computing the probability p_{ij}^k ,
- T_k : number of operations to compute the expected delay for aircraft k ,
- $T(N)$: number of operations to compute the delay statistics of the $D(t)/M(t)/1$ system with N aircraft.

We note that $k \in [1, N]$; and for a fixed k , $i \in [0, k - 2]$; and for a fixed i , $j \in [0, i + 1]$. As noted earlier, p_{ij}^k is computed as a constant multiplied by $p_{i,j-1}^k$ (Specifically, $p_{ij}^k = \frac{i+2-j}{\mu_{k-1}c_{k-1}}p_{i,j-1}^k$) and hence, if one uses a “bottom-up” approach (i.e., p_{ij}^k is computed starting from $j = 0$ to larger values of j), then, $t_{i,j,k} = O(1)$. But, it is important to point out that in the absence of the recursive definition of the probabilities, $t_{i,j,k} \neq O(1)$ as an explicit computation entails the knowledge of a few factorials, which would worsen the overall time complexity of our approach. In the remainder, we shall work with $t_{i,j,k} = O(1)$ (assuming the use of the recursive definition) as this forms an important element of our overall approach.

Next, we show that $T_k = O(k^2)$:

$$T_k = \sum_{i=0}^{k-2} \sum_{j=0}^{i+1} t_{i,j,k} \quad (2.33)$$

$$= \sum_{i=0}^{k-2} \left(\sum_{j=0}^{i+1} O(1) \right) \quad (2.34)$$

$$= \sum_{i=0}^{k-2} ((i+2)O(1)) \quad (2.35)$$

$$= \sum_{i=0}^{k-2} O(i) \quad (2.36)$$

$$= O(k^2) \quad (2.37)$$

Finally, we characterize $T(N)$:

$$T(N) = \sum_{k=1}^N T_k \quad (2.38)$$

$$= \sum_{k=1}^N O(k^2) \quad (2.39)$$

$$= O(N^3) \quad (2.40)$$

The analysis just presented shows that, the time complexity of our proposed numerical method for transient analysis of $D(t)/M(t)/1$ system is polynomial in the total number of aircraft arrivals ($O(N^3)$). One surprising aspect of this analysis is that the time complexity is *independent of the load factor* (ratio of demand rate and service rate) across different time-periods, and only depends on the total number of aircraft arrivals. This can be explained by noting that the number of non-zero p_{ij}^k 's is a function only of the total number of aircraft arrivals (N). This follows because the k^{th} aircraft can only have at most one more aircraft in the queue than those seen by the $(k-1)^{th}$ aircraft. Moreover, the k^{th} aircraft will only have $0, 1, \dots, k-1$ number of aircraft in the queue when it arrives. Hence, the number of non-zero p_{ij}^k 's is a function only of the total number of aircraft arrivals and is independent of the load factor across different time-periods. The load factors can potentially alter the relative magnitude of the various probabilities, but they will not impact the actual number of such non-zero values. Finally, we state that the method is still

computationally tractable when the demand exceeds the capacity for some periods during the day. This is a direct implication of the fact that the time complexity depends only on the number of aircraft arrivals.

Space Complexity.

The space required for our numerical method is a function of the number of non-zero p_{ij}^k 's (transition probabilities in the discrete-time Markov Chain). As noted earlier, the number of non-zero p_{ij}^k 's is a function only of the total number of aircraft arrivals, and it is $O(N^3)$ (as computed during the analysis of time complexity). Hence, the space complexity of our approach is also $O(N^3)$.

2.3.2 Comparison of Steady State Statistics with $M/E_k/1$ System

Here, we compare the basic statistics of a queuing system in steady state (expected number of customers in the queue - \bar{L} and average waiting time - \bar{W}) for a $D/M/1$ system with those of a $M/D/1$ system. We may use the Pollaczek-Khinchine (P-K) formula to compute \bar{W} in steady-state for a $M/G/1$ system ($E[Z]$ and $E[Z^2]$ denote the first and second moments of the service time):

$$\bar{W} = \frac{\lambda E[Z^2]}{2(1 - \lambda E[Z])} \quad (2.41)$$

Further, using Little's law, \bar{L} can be computed as follows:

$$\bar{L} = \frac{\lambda^2 E[Z^2]}{2(1 - \lambda E[Z])} \quad (2.42)$$

Finally, we use our numerical method to compute \bar{W} for a $D/M/1$ system in steady-state (for this we extend the time length of analysis until the system reaches steady-state). We use Little's law to compute \bar{L} , given the value of \bar{W} .

In the entire analysis, we use a service rate of 60 aircraft per hour, which implies that the random variable governing the service time has a mean of 1 minute (i.e., $\frac{1}{\mu} = 1$ minute). With this service rate, we alter the demand rate to capture the various load factors (ρ). In particular, we use a demand rate of 6, 15, 30, 45, 48 and 54 aircraft per hour (leading to $\rho = 0.10, 0.25, 0.50, 0.75, 0.80$ and 0.90 respectively). As a sanity check, we keep in mind that the average delays reported for a $D/M/1$ system must always be lower than those of a $M/M/1$ system.

S.No.	Queuing System	Load Factor (ρ)					
		0.10	0.25	0.50	0.75	0.80	0.90
1	D/M/1	0.100	0.27	0.69	1.78	2.29	4.80
2	M/D/1	0.055	0.16	0.50	1.50	2.00	4.50
3	M/M/1	0.110	0.30	1.00	3.00	4.00	9.00

Table 2.1: Comparison of average waiting time (\bar{W}) in steady-state under different load factors (ρ).

S.No.	Queuing System	Load Factor (ρ)					
		0.10	0.25	0.50	0.75	0.80	0.90
1	D/M/1	0.010	0.068	0.34	1.340	1.83	4.32
2	M/D/1	0.005	0.040	0.25	1.125	1.60	4.05
3	M/M/1	0.011	0.080	0.50	2.250	3.20	8.10

Table 2.2: Comparison of expected number in the queue (\bar{L}) in steady-state under different load factors (ρ).

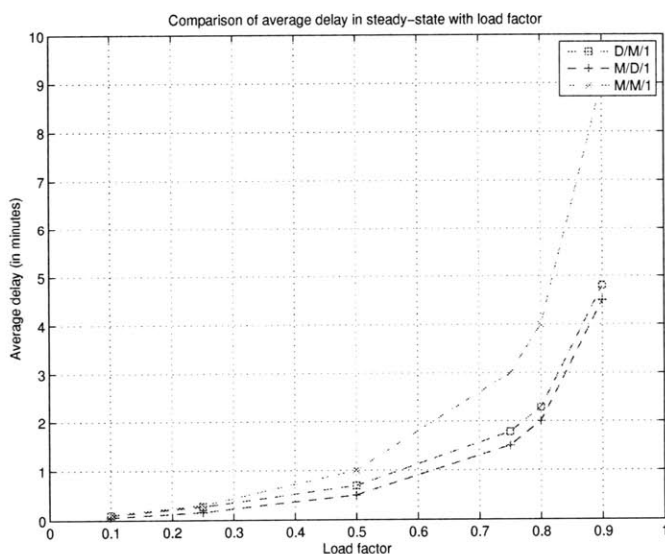


Figure 2.2: Comparison of average delay (\bar{W}) in steady-state with load factor.

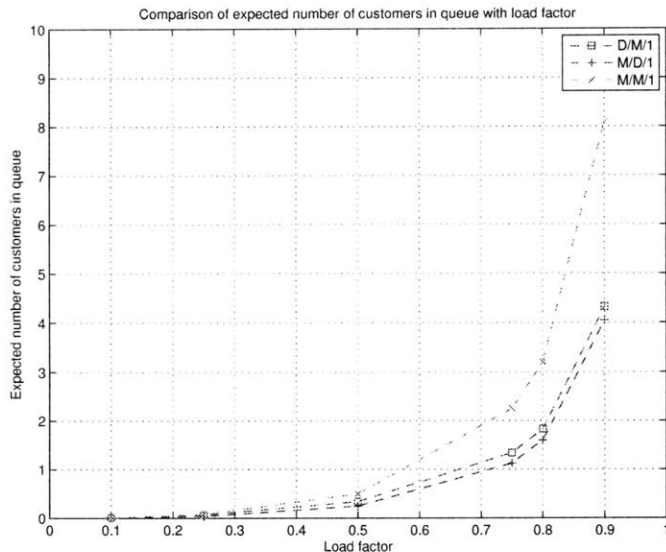


Figure 2.3: Comparison of average number of customer in queue (\bar{L}) in steady-state with load factor.

The numbers reported in Tables 2.1 and 2.2 indicate that in steady-state, both \bar{L} and \bar{W} for a $D/M/1$ system lie between $M/D/1$ and $M/M/1$ respectively. Figures 2.2 and 2.3 plot the graph for the numbers reported in these tables. The key takeaway from these plots is that, as the load factor (ρ) increases, the difference (in “percentage” terms) between $D/M/1$ and $M/D/1$ decreases whereas the opposite is true for the difference between $D/M/1$ values and $M/M/1$ values. Thus, as the level of congestion in the system increases, the $M/D/1$ and $D/M/1$ systems approach closer to each other. (Similar conclusions will also be drawn from the results reported for a set of US airports in Chapter 3 which reinforce these observations). Another striking feature of these results stem from the observation that for a $D/M/1$ system, both \bar{L} and \bar{W} are consistently higher than for $M/D/1$ as ρ increases.

2.3.3 Comparison of Transient Delays with the $M(t)/D(t)/1$ System

In this section, we study the relationship between aggregate transient delays of a $D(t)/M(t)/1$ and $M(t)/D(t)/1$ system as a function of level of congestion in the system and variability of demand over time.

Effect of level of congestion.

Here, we study the relationship between D/M/1 and M/D/1 delays as a function of level of congestion in the system. We consider a test-case where we have a fixed service capacity (60 ac/hr), but three scenarios for demand which vary with respect to level of congestion:

- *Scenario 1*: demand rate is 6 ac/hr (**weakly congested**, $\rho = 0.1$).
- *Scenario 2*: demand rate is 30 ac/hr (**mildly congested**, $\rho = 0.5$).
- *Scenario 3*: demand rate is 54 ac/hr (**strongly congested**, $\rho = 0.9$).

Table 2.3 lists the delays for the D/M/1 and M/D/1 systems under Scenarios 1-3 as listed above. The numbers illustrate two key trends: i) the delays under both systems increase as the level of congestion in the systems increases; and ii) the percentage increase in the D/M/1 delays by comparison to the M/D/1 delays decreases with the increase in system congestion. This implies that the delays associated with each of the two systems approach each other as the level of congestion increases.

S.No.	Demand Rate	Aggregate Delay (in min.)		% Increase in Delays from M/D/1 to D/M/1
		M/D/1	D/M/1	
1	<i>Scenario 1</i>	4	7	75.00
2	<i>Scenario 2</i>	179	250	39.66
3	<i>Scenario 3</i>	2723	3130	14.94

Table 2.3: Effect of level of congestion.

Effect of demand variability over time.

Here, we study the effect of demand variability over time. We consider two different test-cases across a period of 12 hours where we have a fixed capacity (60 ac/hr). In each of these test-cases, we examine three scenarios for demand which are time-varying. Finally, we set a utilization ratio of $\rho = 0.5$ in the first test-case, and $\rho = 0.8$ in the second test-case.

1. Test Case 1 ($\rho = 0.5$)

- *Scenario 4*: demand rate is 30 ac/hr (**not time-varying**).

- *Scenario 5*: demand rate is alternating between 20 and 40 ac/hr (**weakly time-varying**).
- *Scenario 6*: demand rate is alternating between 10 and 50 ac/hr (**strongly time-varying**).

2. Test Case 2 ($\rho = 0.8$)

- *Scenario 7*: demand rate is 48 ac/hr (**not time-varying**).
- *Scenario 8*: demand rate is alternating between 32 and 64 ac/hr (**weakly time-varying**).
- *Scenario 9*: demand rate is alternating between 16 and 80 ac/hr (**strongly time-varying**).

Tables 2.4 and 2.5 report the delays for the D/M/1 and M/D/1 systems under the two test-cases ($\rho = 0.5$ and 0.8) respectively. The reported numbers suggest that the delays under both system increase as the degree of variability in the demand increases. Moreover, the percentage increase in the D/M/1 delays over the M/D/1 delays increase with increases in the degree of demand variability.

S.No.	Demand Rate	Aggregate Delay (in min.)		% Increase in Delays from M/D/1 to D/M/1
		M/D/1	D/M/1	
1	<i>Scenario 4</i>	179	250	39.66
2	<i>Scenario 5</i>	207	329	58.93
3	<i>Scenario 6</i>	302	592	96.02

Table 2.4: Effect of demand variability over time (Test Case 1: $\rho = 0.5$).

S.No.	Demand Rate	Aggregate Delay (in min.)		% Increase in Delays from M/D/1 to D/M/1
		M/D/1	D/M/1	
1	<i>Scenario 7</i>	1134	1343	18.43
2	<i>Scenario 8</i>	1662	2525	51.92
3	<i>Scenario 9</i>	3102	6021	94.10

Table 2.5: Effect of demand variability over time (Test Case 2: $\rho = 0.8$).

An important conclusion from the results reported in Subsections 2.3.2 and 2.3.3 is that M/D/1 and D/M/1 give increasingly close results (percent-wise) as the system

utilization increases (whether with constant or time-varying demand and capacity profiles); and, for a given system utilization, the percent-wise difference increases as variability with time increases.

2.4 Summary

In this chapter, we develop a numerical method for the transient analysis of a $D(t)/M(t)/1$ system for both constant and time-varying service rates. The analysis relies on computing the probabilities of the queue length at various aircraft arrival epochs. Moreover, we propose a recursive way to compute the state probabilities to make the procedure tractable and propose the concept of an *effective service rate* to process the queue. We conclude by presenting an analysis of the complexity (time and space) of our approach; a comparison of the steady-state and transient delays of this queuing system with the $M(t)/D(t)/1$ system; and the effect of level of congestion and time variation of demand on system delays.

Chapter 3

Comparing Airport Delay Estimates from Various Queuing Systems

In this chapter, we use a set of models of dynamic queuing systems, in addition to our $D(t)/M(t)/1$ model, to compute airport delays under the different queuing systems in a dynamic setting. Specifically, we focus on a set of six major US airports for which we have data on the actual demand and capacity profiles. Section 3.1 presents a brief description of the queuing models (in addition to the $D(t)/M(t)/1$ model) used in our comparison. Section 3.2 reports the empirical results from the various queuing models on the six airports. Section 3.3 presents some additional comparisons on the reported results.

3.1 Other Queuing Systems

In this section, we present a brief description of the additional queuing systems ($M(t)/E_k(t)/1$ and $D(t)/D(t)/1$) and the associated computing tools that can be used to obtain approximate numerical solutions (namely, DELAYS, Micro and Macro models). These solutions will be used to compare airport delays.

3.1.1 The $M(t)/E_k(t)/1$ System

The $M(t)/E_k(t)/1$ system in queuing theory denotes a model with a non-stationary Poisson arrival process, time-dependent k^{th} -order Erlang service-time distribution,

a single-server, and infinite waiting room. In the context of this thesis, this system can be used to compute statistics for $M/D/1$ for k large and $M/M/1$ for $k = 1$. These two systems comprise a subset of the four queuing systems that we envisioned in the Introduction to characterize the different levels of uncertainty modeling.

Given a demand rate $\lambda(t)$ and a service rate $\mu(t)$, a set of first-order differential equations (often referred to as the ‘‘Chapman-Kolmogorov equations’’) describe the evolution over time of the $M(t)/E_k(t)/1$ queuing system, when the queue capacity is equal to N customers. The typical steady-state diagram of this queuing system consists of $kN + 1$ ‘‘stages’’. In order to approximate an infinite-capacity system, the queue capacity, N , of the system must be sufficiently large so that the probability that the system is full at any time is very small. Unfortunately, for $M(t)/E_k(t)/1$ systems, the number of equations to be solved can become very large quickly (for large k , i.e., when the service times have a small coefficient of variation or, in practical terms, are nearly constant), and so the numerical solution of the system is time-consuming.

For this reason, we use a tool developed by Kivestu (1976) at MIT (called DELAYS) that provides an approximate approach to the $M(t)/E_k(t)/1$ system based on solving a set of $N + 1$ difference equations (independent of k), instead of a system of $kN + 1$ equations. The Chapman-Kolmogorov (C-K) first-order differential equations that describe the evolution over time of a $M(t)/E_k(t)/1$ queuing system with a capacity of N customers (i.e., queuing space for $N - 1$ customers) are as follows:

$$P_0(t) = -\lambda(t)P_0(t) + k\mu(t)P_1(t), \quad (3.1)$$

$$P_i(t) = -(\lambda(t) + k\mu(t))P_i(t) + k\mu(t)P_{i+1}(t), \quad 1 < i < k - 1, \quad (3.2)$$

$$P_i(t) = \lambda(t)P_{i-k}(t) - (\lambda(t) + k\mu(t))P_i(t) + k\mu(t)P_{i+1}(t), \quad k < i < (N - 1)k, \quad (3.3)$$

$$P_i(t) = \lambda(t)P_{i-k}(t) + k\mu(t)(P_{i+1}(t) - P_i(t)), \quad (N - 1)k < i < Nk - 1, \quad (3.4)$$

$$P_{Nk}(t) = -k\mu(t)P_{Nk}(t) + \lambda(t)P_{(N-1)k}(t). \quad (3.5)$$

where $\lambda(t)$ is the demand rate, $\mu(t)$ is the service rate, i is the current state of the queuing system, and $P_i(t)$ is the probability of being in state i at time t .

In airport applications, DELAYS uses the number of landing and/or takeoff demands to set the demand rate, $\lambda(t)$, and the airport’s capacity profile to set the service rate, $\mu(t)$. DELAYS then starts at time $t = 0$ (when the airport can be

said to be “at rest”) with $P_0(0) = 1$ and $P_i(0) = 0$ for all $i > 0$. The system is assumed initially to be empty (this matches well with the NAS) and has negligible probability of returning to an idle state throughout the analysis period. The model then proceeds to solve the C-K equations iteratively by computing $P_i(\Delta t) = 0$ for all system states $i(0 \leq i \leq Nk)$ and for an appropriately small time increment, Δt , and continues by computing $P_i(2\Delta t)$, $P_i(3\Delta t)$ etc., for all system states i , until it has “stepped” in this way through a 24-hour period. Having thus computed all the state probabilities for the entire 24-hour period, it then provides estimates for the average waiting times, average number of aircraft in the queue for using the runway system, etc., as specified by the user.

Because DELAYS is an approximation of a queuing system with infinite queuing space, the number of states for which the C-K equations are solved is set to a large number from the outset. Moreover, at the end of every iteration, DELAYS checks the probability of having a full queuing system. If that probability exceeds 10^{-6} , then DELAYS adds k more states to the already existing ones. In this way, the system always has sufficient queuing space to accommodate all incoming demands. With the time parameter Δt set small enough, any potential state space violation can be remedied before it affects the evolution of the system.

During Fall 2007, this author rewrote the code in Java. It runs very efficiently on any typical PC (less than 1 sec for a 24-hour period at the busiest airports). Moreover, this author has also proposed a variant of DELAYS (called DELAYS-EFF and described later in this chapter), which uses the concept of *effective service rate* instead of instantaneous service rate to process the queue and rectify the rare cases where DELAYS overestimates/underestimates the actual delays due to service rates that vary drastically over time.

3.1.2 Two Versions of a $D(t)/D(t)/1$ System

The delays in a completely deterministic setting are computed using two models, Micro and Macro, that were developed at the University of California, Berkeley. The Micro model uses as its demand input the scheduled times of arrivals and/or departures at an airport, while the Macro model is based on aggregate flight demands in discrete time periods. Below, we provide a brief description of these two models (please refer to [10] for detailed descriptions).

Micro model.

The Micro model uses individual flight data in order to calculate delays for a deterministic queuing analysis. Specifically, the scheduled times of flights (arrivals or departures) are used as the time when deterministic demands to the system occur. Further, the capacity information during each discrete time-period (referred to as a bin henceforth) is used to compute the minimum headway requirement between two consecutive arrivals. For a particular flight, an expected time of service is assigned, given the information on its scheduled arrival time and the minimum separation required from the immediately previous serviced flight. This is done in a recursive way, simply by noting the expected service time of the just serviced flight and the headway required during the current bin. Finally, the delay for each flight is calculated as the difference between the expected time of service and the scheduled arrival time.

Macro model.

In contrast to the Micro model, the Macro model uses aggregate flight data to determine arrival demand. The capacity for each bin is estimated in a way similar to the method used by the Micro model. Therefore, the difference between the Micro and Macro model lies in the way demand is handled. Given the demand and capacity data for each bin, a cumulative demand curve and throughput curve are created which capture the aggregate number of demands and the aggregate number of serviced aircraft, respectively, at the end of a bin. The delays (for a particular bin) are then computed as the product of the length of the bin and the difference in the cumulative demand and throughput at the end of this bin.

This discussion of the two deterministic models indicates that the Macro model does not capture the unevenness of the true schedule (in contrast to Micro), and hence, is expected to predict lower delays than Micro.

3.2 Delays Under Different Queuing Systems

In this section, we report extensive computational results for a set of US airports for the purpose of comparing the delay estimates provided by four different queuing systems.

Experimental Setup.

We consider six major airports of the US - Atlanta Hartsfield International (ATL), Boston Logan International (BOS), New York LaGuardia (LGA), Miami International (MIA), Chicago O'Hare (ORD) and San Francisco International (SFO) – over a 17-hour period. For each airport, we have access to the demand and capacity profile over the 17-hour time horizon. Specifically, the entire time horizon is typically divided into time-bins (called bin henceforth) of either 15 minutes or 60 minutes and, demand is specified as the number of aircraft arrivals demanding service in each of these bins. The demand profiles were provided by Michel Santos from Intelligent Automation Inc (IAI). The demand schedules are obtained through runs of the Airspace Concept Evaluation System (ACES) - a high-fidelity simulation model of the National Airspace System (NAS). The schedules are based on simulation runs for four peak and four off-peak days in 2007, a total of eight demand scenarios. The output from each of these simulations includes landing/wheels-down day and time in UTC time zone, arrival day and time in UTC time-zone and arrival airport code. From this data set, demand at any airport can be extracted since the simulation is performed for the entire NAS. Moreover, the capacity of individual airports is based on airport acceptance rate (AAR) values obtained from the Airspace System Performance Metric (ASPM) data base maintained by the Federal Aviation Administration (FAA). ASPM provides quarter-hour AAR values for every day beginning on January 1, 2000. The capacity profiles from this information were prepared by Prof. Hansen and his research team at Berkeley. For this, a statistical analysis of capacities available during the 365 days of 2007 was carried out, and five “clusters” of representative profiles were filtered. For instance, one cluster might represent a day when weather was good throughout the day, another a day when the weather deteriorated during the day, so it was good in the morning and poor later on, etc. Each cluster is assigned a weight, based on the frequency with which the associated conditions occur.

Given these demand and capacity datasets, the delays arising from deterministic settings are computed using the Berkeley Micro and Macro models. Next, we compute delays resulting from adding stochasticities in the arrival and service process. DELAYS makes possible the computation of delays in a $M/E_k/1$ system, where k is the Erlang parameter for the order k Erlang service time distribution (which is a sum of k i.i.d exponential random variables). We report results for $k = 100000, 10$ and 1

which correspond to $M/D/1$, $M/E_{10}/1$ and $M/M/1$ systems respectively. Finally, we use the analysis presented in Chapter 2 to compute delays in a $D/M/1$ system. We run our experiments on a laptop with 2 GB RAM and Linux Ubuntu OS. DELAYS runs very efficiently taking less than 1 sec at the busiest airports, whereas, the code for $D/M/1$ system typically takes less than 10 minutes for the most congested airports.

Next, we report the delays experienced at all the airports (except LGA) mentioned above under the four dynamic queuing systems. Tables 3.1-3.5 report the delays for these airports under the different capacity scenarios. Figures 3.1-3.6 plot the capacity profiles used for computing delays at these airports.

Table 3.6 lists the delays for the $D/M/1$ and $M/D/1$ systems (averaged over the different capacity profiles). As is evident from the results, the delays under a $D/M/1$ system are higher than those for the $M/D/1$ system in all cases. But, the percentage increase in the average delays of a $D/M/1$ system compared to a $M/D/1$ system decreases as the system becomes more congested. In fact, for the two most congested airports (ATL and SFO), the average delays are practically the same.

S.No.	Queuing System	Capacity Scenario				
		ATL1	ATL2	ATL3	ATL4	ATL5
1	D/D/1 (Macro)	3577	2104	8385	11773	30714
2	D/D/1 (Micro)	4861	3336	9166	12412	31257
3	M/D/1	6525	4279	12460	14550	35475
4	D/M/1	6878	4620	12815	14901	35954
5	M/E ₁₀ /1	6805	4469	12836	14897	35982
6	M/M/1	9092	6099	15938	17795	40154

Table 3.1: Comparison of aggregate delays for a 17-hour period (in min) under different queuing systems - ATL.

Interpretation and takeaways from computations.

Next, we draw some inferences from the results tabulated in Tables 3.1-3.6:

Range of delay values.

The results indicate that the delays resulting from the queuing systems can be broadly divided into three distinct bands. The delays from $M/D/1$, $D/M/1$ and $M/E_{10}/1$ systems lie close to each other within a narrow band. The, $D/D/1$ and $M/M/1$ systems hold the other two end-points on opposite sides of this band. The

S.No.	Queuing System	Capacity Scenario				
		BOS1	BOS2	BOS3	BOS4	BOS5
1	D/D/1 (Macro)	2006	66	241	642	1640
2	D/D/1 (Micro)	3029	372	793	1473	2234
3	M/D/1	6173	454	1133	2730	4046
4	D/M/1	6428	594	1327	2956	4295
5	M/E ₁₀ /1	6476	492	1219	2913	4242
6	M/M/1	8949	829	1945	4434	5823

Table 3.2: Comparison of aggregate delays for a 17-hour period (in min) under different queuing systems - BOS.

S.No.	Queuing System	Capacity Scenario			
		MIA1	MIA2	MIA3	MIA4
1	D/D/1 (Macro)	49	29	98	65
2	D/D/1 (Micro)	174	145	281	283
3	M/D/1	274	239	424	478
4	D/M/1	383	340	557	617
5	M/E ₁₀ /1	298	261	458	516
6	M/M/1	509	450	753	843

Table 3.3: Comparison of aggregate delays for a 17-hour period (in min) under different queuing systems - MIA.

S.No.	Queuing System	Capacity Scenario				
		ORD1	ORD2	ORD3	ORD4	ORD5
1	D/D/1 (Macro)	33	113	50	367	116
2	D/D/1 (Micro)	431	639	616	1144	623
3	M/D/1	445	705	686	1381	744
4	D/M/1	594	900	867	1640	951
5	M/E ₁₀ /1	485	764	744	1479	804
6	M/M/1	833	1265	1235	2312	1322

Table 3.4: Comparison of aggregate delays for a 17-hour period (in min) under different queuing systems - ORD.

S.No.	Queuing System	Capacity Scenario				
		SFO1	SFO2	SFO3	SFO4	SFO5
1	D/D/1 (Macro)	3182	17	8058	92	367
2	D/D/1 (Micro)	3823	220	9200	362	800
3	M/D/1	5281	355	14547	610	1269
4	D/M/1	5305	468	14764	758	1446
5	M/E ₁₀ /1	5422	387	14991	659	1349
6	M/M/1	6624	658	18205	1079	2045

Table 3.5: Comparison of aggregate delays for a 17-hour period (in min) under different queuing systems - SFO.

S.No.	Airport	Average Delay (in minutes)				% Increase in Delays from M/D/1 to D/M/1
		Aggregate		Per flight		
		M/D/1	D/M/1	M/D/1	D/M/1	
1	ATL	14657.8	15033.6	12.57	12.89	2.50
2	BOS	2907.2	3120	5.36	5.76	7.32
3	MIA	353.75	474.25	0.79	1.07	34.06
4	ORD	792.2	990	0.95	1.18	25.01
5	SFO	4412.4	4548.2	9.32	9.61	3.07

Table 3.6: Comparison of Delays (averaged over all capacity scenarios) computed for the D/M/1 and M/D/1 Systems.

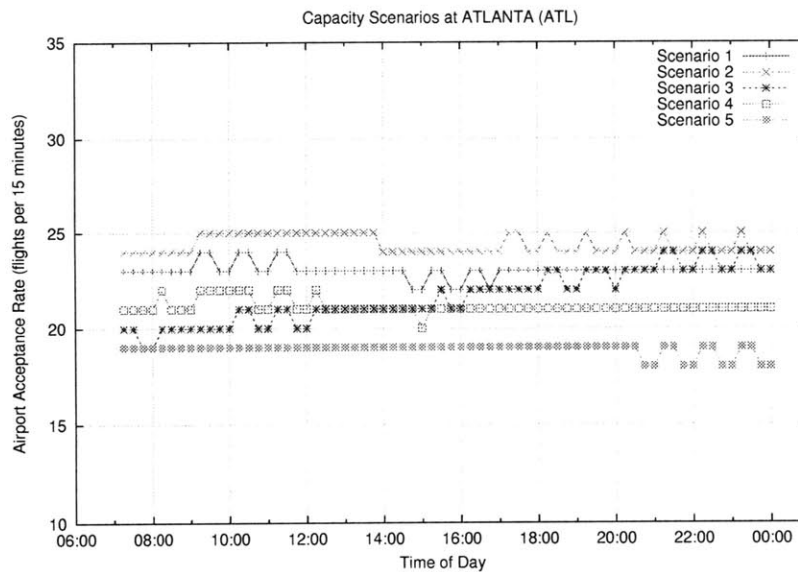


Figure 3.1: Capacity Scenarios at ATL.

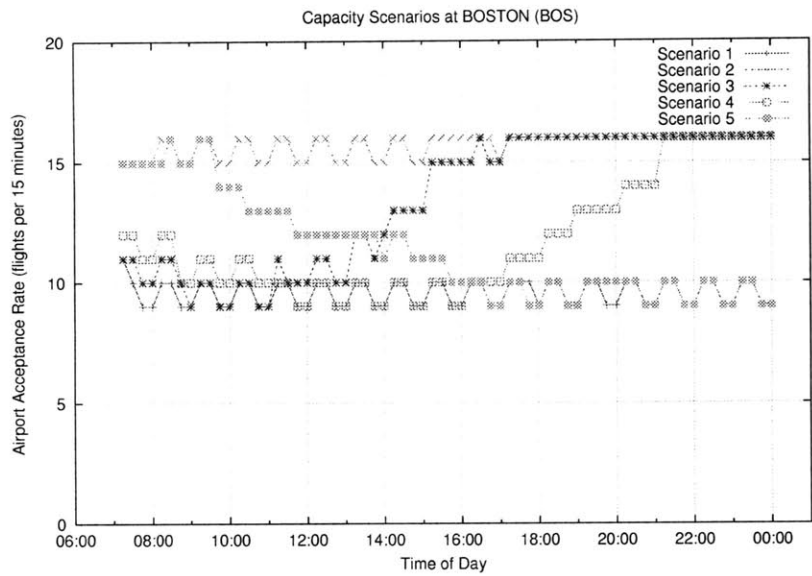


Figure 3.2: Capacity Scenarios at BOS.

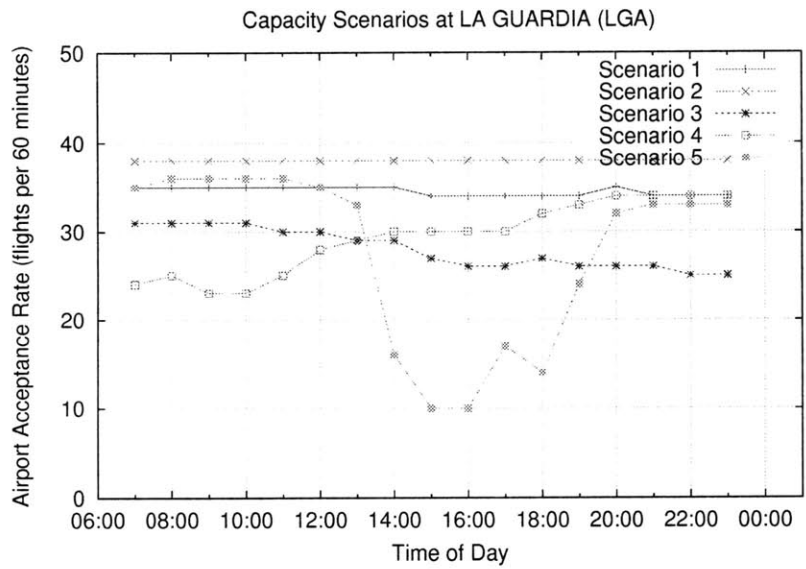


Figure 3.3: Capacity Scenarios at LGA.

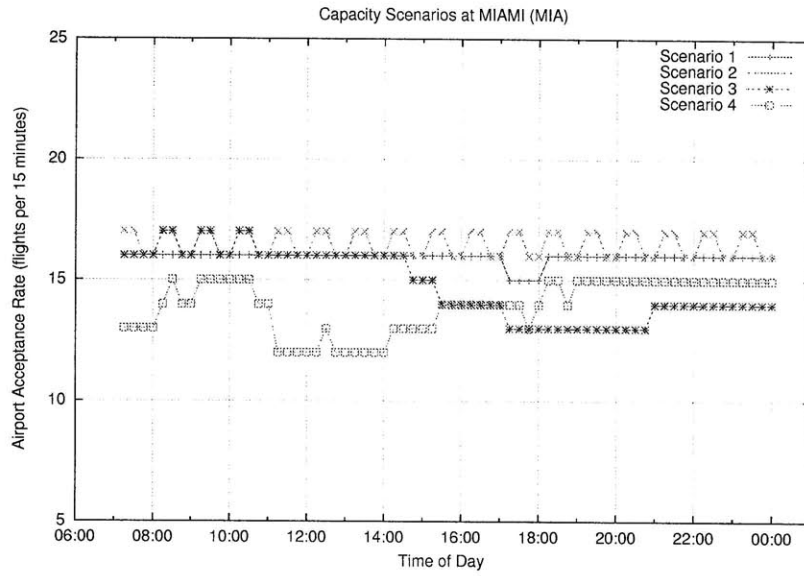


Figure 3.4: Capacity Scenarios at MIA.

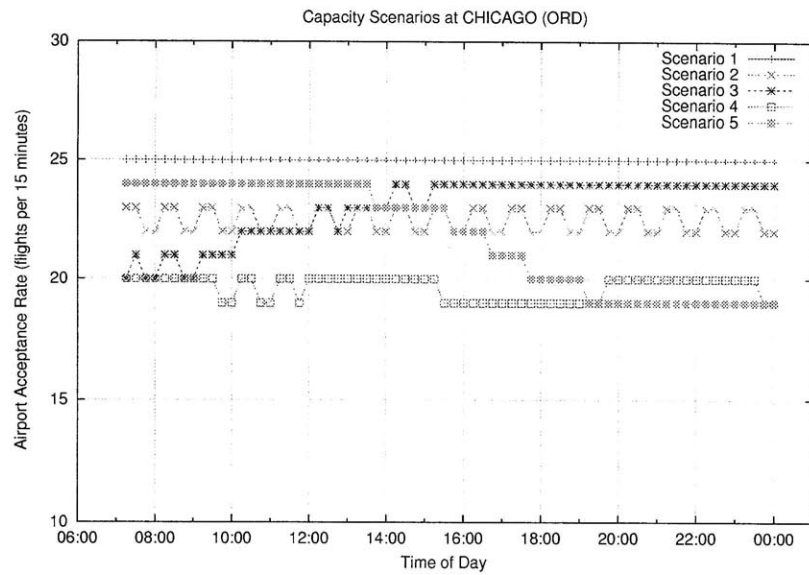


Figure 3.5: Capacity Scenarios at ORD.

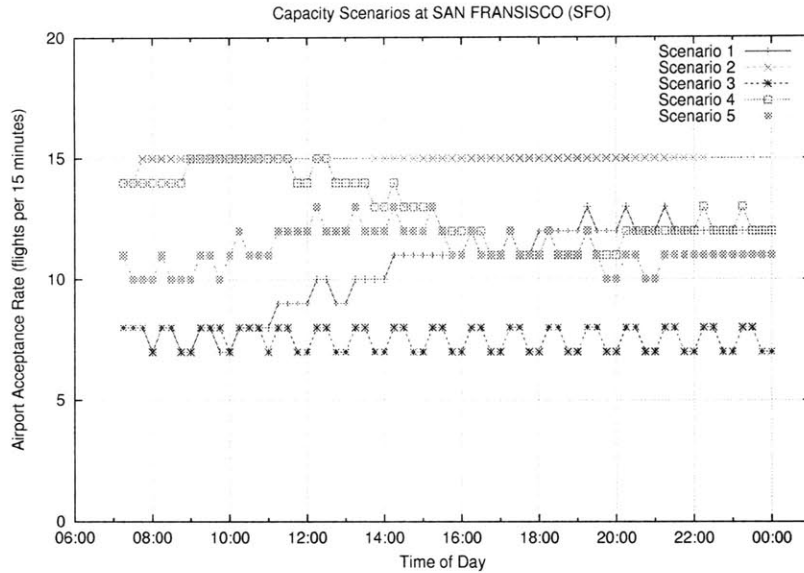


Figure 3.6: Capacity Scenarios at SFO.

wide range of estimates obtained indicate that, improvements in either the predictability of the arrival process or the service process may lead to very substantial reductions in delays. Further, a transition from the intermediate precision cases to "ideal" deterministic scenarios (which, of course, may be extremely difficult to achieve in practice) will lead to another level of improvement in system delays.

Relation between D/M/1 and M/D/1 delays.

Another key takeaway from the results is that the fit between the results of the $D/M/1$ and $M/D/1$ models is very good with a R^2 (goodness-of-fit) measure of 0.99 (see Figure 3.7). The delay estimates provided by these two queuing systems are best compared by dividing the set of airports into two groups - i) congested (ATL, BOS and SFO); and ii) non-congested (MIA and ORD). At the two congested airports, since the average per flight delays are greater than 5 minutes (see Table 3.6), the intercept of -0.27 in the regression line has a small effect and hence, the slope of 0.99 suggests that the increase in $D/M/1$ delays over a $M/D/1$ system is typically of the order of 1%. In contrast, the per flight delays for non-congested airports are typically less than two minutes and, hence, the intercept has a significant impact leading to a higher percentage increase in the delays estimated by $D/M/1$ system over a $M/D/1$ system (also corroborated by Table 3.6). In summary, the results indicate

that in all scenarios, the delay estimates under a $D/M/1$ system are higher than under the $M/D/1$ system and the percentage increase is much higher for non-congested airports, whereas for congested airports, the delay estimates are practically the same.

Results from Macro model.

For some of the least congested days (e.g., BOS2, ORD1, ORD3, SFO2 and all cases of MIA), the difference between Macro and $D/M/1$ or $M/D/1$ (let alone $M/M/1$) is 1-to-5 or more. In addition, the difference between Macro and Micro can be very large in such cases, pointing out the lack of usefulness of the Macro model for such cases. The reason is that Macro does not use any information on the actual arrival times, while only utilizing the aggregate demand during each bin. Hence, this characteristic of the Macro model coupled with the small demand compared to capacity for uncongested days leads to relatively miniscule delay estimates. In contrast, the Micro model is able to reasonably overcome this shortcoming by taking into account the unevenness associated with the actual times of deterministic arrivals.

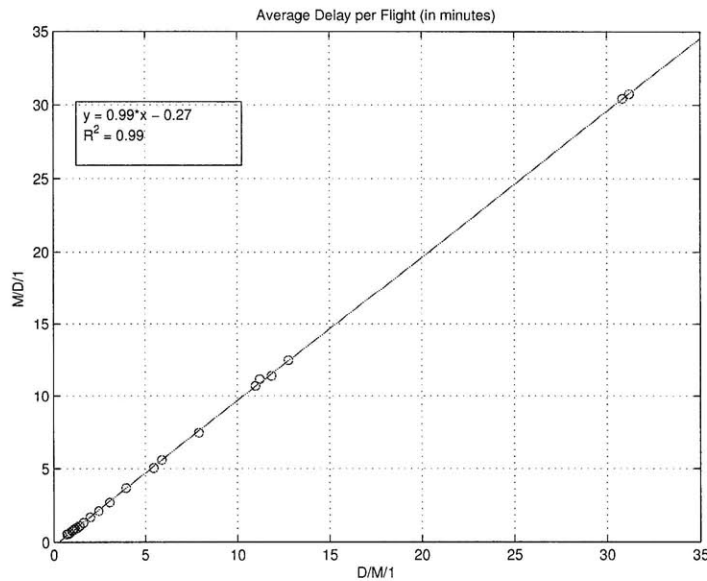


Figure 3.7: Regression Between $D/M/1$ and $M/D/1$ Queuing Systems.

3.3 Additional Comparisons

In this section, we provide further comparisons of the delay estimates obtained through the following three models - Stochastic (DELAYS), Deterministic (Micro) and Deterministic (Macro). Figures 3.8, 3.9 and 3.10 compare the average delay estimated by the Micro and Macro models, the DELAYS and Micro models, and the DELAYS and Macro models respectively, for the various scenarios examined at the subject airports.

It is evident from Figure 3.8 that the delay estimates from the Macro model are slightly, but consistently, below those from the Micro model. Such an observation is expected because the Macro model uses a smoothed demand profile as compared to Micro which takes into account the exact scheduled arrival times of flights, which may be far from evenly spread over time at many major airports. A R^2 (goodness-of-fit) measure of 0.99 indicates that the linear correlation is very strong.

Comparisons of the delay estimates obtained from the DELAYS model with those from the deterministic models are less straightforward. Specifically, there appear to be a set of six points at the extreme right in Figures 3.9 and 3.10 with unreasonably high magnitude of delay estimates (greater than two hours). These points correspond to a specific capacity scenario of LGA (Scenario 5). We elaborate on this unusual behavior in Section 3.3.2. For now, we focus on the comparison of the delay estimates obtained across all points except these six. For this set of points, the best fit line between DELAYS and Micro has a slope of 0.90 and an intercept of -0.84, i.e., given a DELAYS estimate, there is a reduction of 10% and a subsequent reduction of 0.84 minutes in computing the Micro estimate. A similar analysis between DELAYS and Macro shows a reduction of 12%, plus 1.5 minutes.

3.3.1 Effect of Congestion

The plots in Figures 3.9 and 3.10 indicate that there is some correlation amongst the delays obtained from a stochastic model (DELAYS) and a deterministic model (Micro or Macro) if one examines the respective results of these models as a function of level of congestion. Hence, in this section, we try to provide a qualitative explanation of this phenomenon:

1. *Low Congestion:* Here, we would expect that DELAYS would provide higher delay estimates than the deterministic models, possibly much higher on a percentage basis. In fact, in those cases where the demand rate is significantly

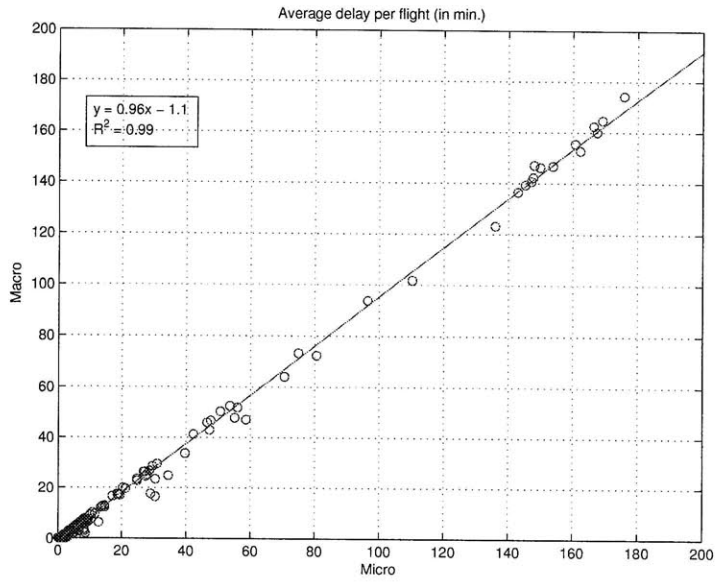


Figure 3.8: Regression between Micro and Macro.

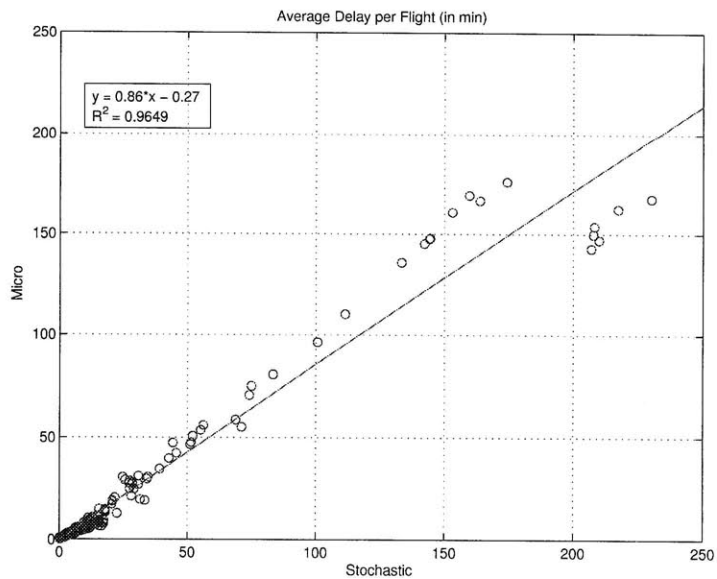


Figure 3.9: Regression between DELAYS and Micro.

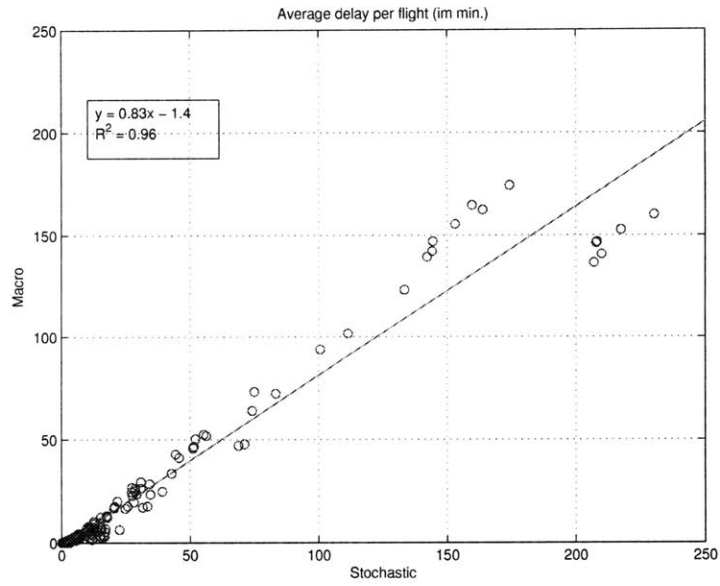


Figure 3.10: Regression between DELAYS and Macro.

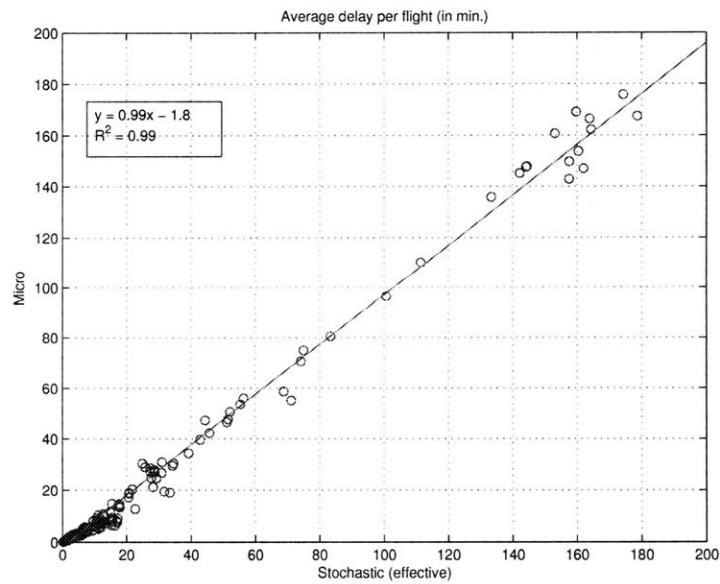


Figure 3.11: Regression between DELAYS-EFF and Micro.

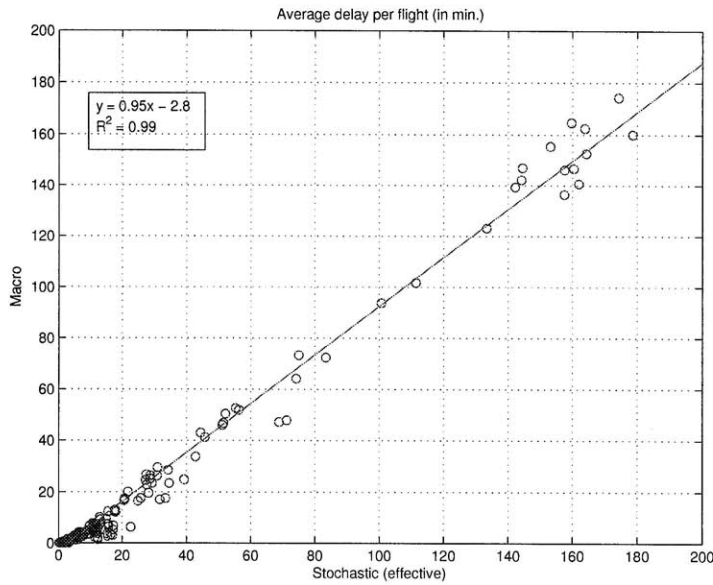


Figure 3.12: Regression between DELAYS-EFF and Macro.

lower than the service rate at all times during the day, a deterministic model might produce a small delay estimate (possibly close to zero), whereas DELAYS might produce an estimate that can be a large number if we set the demand rate to be only slightly below the service rate for long periods of time during the day. The points on the extreme lower left part of the two graphs (i.e., for average delays of 20 minutes or less) corroborate this reasoning.

2. *High Congestion:* Here, we would expect that the results of DELAYS and of the deterministic model would be quite similar, since the delays are dominated by the fact that the demand rate exceeds the service rate for much of the time and, thus, the effects of stochasticity are less important. This is also partly borne out by the graphs where points fall close to the 45° line as we move toward the right.

In summary, the level of congestion at an airport appears to have a pronounced impact on the closeness of the delay estimates obtained from DELAYS, Micro and Macro models.

3.3.2 Some Anomalous Cases

In contrast to the trends observed in Section 3.3.1, we highlight next some anomalous cases in Figures 3.9 and 3.10. Specifically, i) we focus on a scenario where the Micro model produces a higher delay estimate than DELAYS; and ii) the set of six points (LGA cases) which exhibit unusual behavior in Figures 3.9 and 3.10.

Point above 45° line.

There is a single point in the graph which falls above the 45° line (not by much, but nonetheless higher). This means that for this point, the delay estimates provided by Micro are higher than DELAYS. At first glance, this might seem impossible, as a stochastic model should always lead to a higher delay estimate than a deterministic model. We explain below how such a paradoxical situation is possible in an airport setting.

As mentioned earlier, the Micro model uses the scheduled times of flights (arrivals or departures) as the time when the deterministic demands to the system occur. This means that, if the scheduled times are bunched together, many demands will occur within a short span of time and a long queue may be formed. This is often exactly what happens at some of the hub airports, where airlines, for competitive reasons, schedule many flights to depart on the hour or on the half-hour, and also schedule many flights to arrive within a short span of time, so that their passengers can alight from their incoming flight and transfer to a departing flight. (These are the so-called “airline banks” or “airline waves”). It is therefore possible that the demand stream for the Micro model will have more “clusters” of demands than a Poisson arrival stream, for which demands are spread independently and randomly over time. Thus, there may be a few cases in which delays under the Micro $D(t)/D(t)/1$ model could be greater than delays under a stochastic model.

The case in which the Micro model produces greater delay than the stochastic model (above 45-degree line) involves Chicago O’Hare (ORD) Airport. ORD, a major hub for both American and United airlines, is notorious for the bunching of its flights into “banks” or “waves”.

LGA cases.

There is a distinct bunch of six points in Figures 3.9 and 3.10 which appear outliers given the general trend. These cases correspond to a particular capacity scenario

of LGA (Scenario 5). These confounding outcomes can be traced back to the exact mechanics of the DELAYS model. DELAYS works by first dividing the entire time-horizon of analysis into several time-bins (this division is a function of the Erlang parameter k of the service process). Subsequent to this, it approximates the aircraft delays accrued in each of these bins by using the instantaneous service rate during that particular bin to process the queue of aircraft. This, in turn, means that, if the service rate changes drastically in the near future, starting from this bin, a poor estimate of the actual delays will result (an over-estimate if the service rate later increases, and an under-estimate if it decreases). For the LGA cases, this is exactly the reason why DELAYS produces very high estimates as there is a drop in the capacity at later in during the day which causes DELAYS to overestimate the accrued delays in the middle of the day.

To rectify this issue, we utilize the concept of *effective service rate* as introduced during the analysis of the $D/M/1$ system, and develop a variant of DELAYS (called DELAYS-EFF henceforth), which essentially replaces the instantaneous service rate used during each time-bin with an effective rate computed in a manner entirely analogous to the one described in Chapter 2 [expression (2.30)]. It should be noted that if the service rate does not vary significantly during the day, DELAYS and DELAYS-EFF will produce practically same results. It is important to comment that there is no rigorous theoretical justification for the correctness of DELAYS-EFF, and a simulation would serve as a good way to confirm the correctness of the resulting delay estimates. Figure 3.11 plots the average delays obtained from Deterministic (Micro) and the DELAYS-EFF model, while Figure 3.12 plots the average delays obtained from Deterministic (Macro) and the DELAYS-EFF model. As expected, the points corresponding to the LGA cases become more aligned with the general trend.

Next, we examine closely the detailed profiles of demand, capacity and delays estimated by the DELAYS model over a 17-hour period for one of outlier points corresponding to the LGA cases (Figure 3.13). A close inspection of the plot indicates that between 14:00 and 18:00, the capacity is extremely low, so, based on the explanation of the mechanics of the DELAYS model (and the use of instantaneous service rates therein), the model predicts per flight delays of 400 minutes (about 7 hours)! After 18:00, the capacity recovers and estimated delays gradually taper off. Overall, the model ends up overestimating delays. In practice, such cases do not occur, because the airlines will cancel numerous flights when the FAA announces

such a drastic drop in capacity later on in the day.

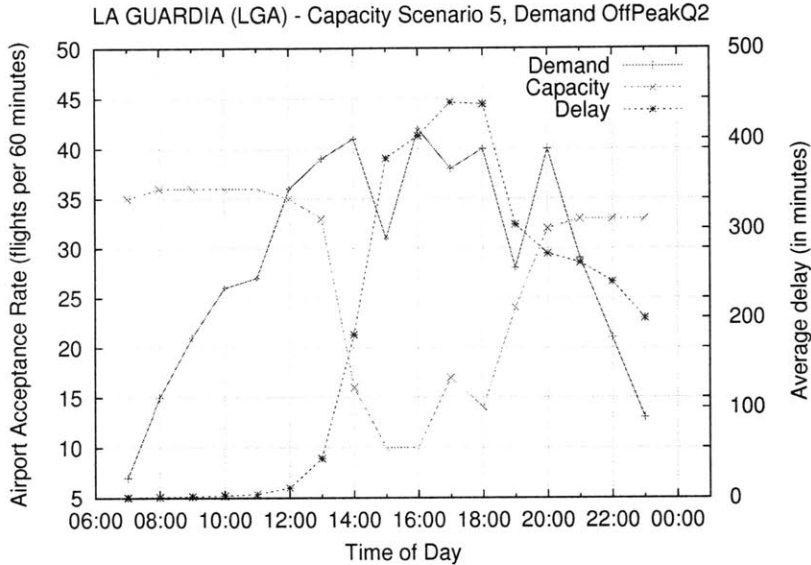


Figure 3.13: One of the outlier points (LGA).

3.3.3 Regression as a Function of Level of Congestion.

Next, we examine the best-fit linear relations of the per flight average delays estimated by the DELAYS-EFF, Micro and Macro models as a function of congestion. As noted earlier, there is a strong correlation between the stochastic and deterministic delays as a function of the level of congestion in the system. A linear regression applied to the average delays computed in all cases, without consideration of the level of congestion present, has the shortcoming of ignoring this important characteristic. Therefore, we perform separate linear regressions for the low and high congestion cases and this analysis is tabulated in Table 3.7. We distinguish the level of congestion into two cases: i) *Low Congestion* (per flight average delay ≤ 20 minutes); and ii) *High Congestion* (per flight average delay > 20 minutes). Table 3.7 indicates that for the low congestion cases, the Micro and Macro produce delay estimates which are smaller than those of the DELAYS-EFF model by 36% and 46%, respectively. For the high congestion cases, Micro and Macro generate estimates which differ from those of the DELAYS-EFF model only in the intercept, 3.7 and 6.4 minutes less, respectively.

S.No.	Scenario	x-axis	y-axis	Equation of Best linear-fit	R^2
1	<i>Low Congestion</i>	DELAYS-EFF	Micro	$y = 0.64x - 0.01$	0.98
2	<i>High Congestion</i>	DELAYS-EFF	Micro	$y = x - 3.7$	0.99
3	<i>Low Congestion</i>	DELAYS-EFF	Macro	$y = 0.54x - 0.52$	0.95
4	<i>High Congestion</i>	DELAYS-EFF	Macro	$y = 0.99x - 6.4$	0.99

Table 3.7: Regression analysis as a function of level of congestion.

3.4 Summary

In this chapter, we provide a brief description of the queuing models (in addition to the $D(t)/M(t)/1$ model) used for a comparison of delay estimates for a set of major and often congested airports in the United States. The key conclusions from the computational results are that the range of resulting delay estimates can be broadly divided into three distinct bands - namely, deterministic ($D(t)/D(t)/1$), intermediate precision ($D(t)/M(t)/1$ and $M(t)/D(t)/1$) and completely stochastic ($M(t)/M(t)/1$); the level of congestion at an airport has a strong impact on the closeness of the delay estimates obtained from the various models; and $D(t)/M(t)/1$ always results in higher delays than a $M(t)/D(t)/1$ system.

Chapter 4

Conclusion

In this thesis, we develop the capability of solving a set of dynamic queuing systems with the aim of comparing airport delays. These queuing systems capture the entire range of uncertainties expected during the deployment of various NGATS technologies. In particular, we present an efficient approach for the computation of delays in a dynamic $D(t)/M(t)/1$ queuing system. In addition, we compare the delays resulting from this system with $D(t)/D(t)/1$ and $M(t)/E_k(t)/1$ (specifically for $k = 1$ and $k = 10000$).

Our results indicate that the closeness of the delay estimates provided by the different models depends on the level of congestion at an airport and the relative shapes of the dynamic profiles of capacity and demand at the airport; the difference (on a “percentage” basis) between the estimates provided by the deterministic model and the stochastic ones is largest for uncongested airports and decreases as the level of congestion increases. $D(t)/M(t)/1$ and $M(t)/D(t)/1$ produce estimates of the same order of magnitude, and reflect delays in the presence of “moderate” uncertainty at an airport; and delays under a $D(t)/M(t)/1$ queuing system are always higher than under a $M(t)/D(t)/1$ system.

Bibliography

- [1] A. Odoni, E. Roth (1983). Empirical Investigation of the transient behaviour of stationary queuing systems. *Operations Research*, 31:432-455.
- [2] B. Koopman (1972). Air Terminal Queues Under Time-Dependent Conditions. *Operations Research*, 20:1089-1114.
- [3] D. Daley. The correlation structure of the output of some single server queuing systems, *Ann. Math. Statist.*, 39, pp. 1007-1019.
- [4] J. Birger (1966). Choosing a Good Appointment System—Study of Queues of the Type (D, M, 1). *Operations Research*, Vol. 14, No. 2, March-April 1966, pp. 292-312.
- [5] Joint Planning and Development Office, Concept of operations for the Next Generation Air Transportation System, Version 2.0, 2006.
- [6] K. Malone (1993). Modeling a Network of Queues Under Nonstationary and Stochastic Conditions. Master's Thesis, MIT.
- [7] L. Green and P. Kolesar (1991). The Pointwise Stationary Approximation for Queues with Nonstationary arrivals. *Management Science*, 1991.
- [8] M. Peterson, D. Bertsimas, A. Odoni (1995). Models and Algorithms for Transient queuing Congestion at Airports. *Management Science*, 1995.
- [9] M. Hansen, T. Nikoleris, D. Lovell, K. Vlachou, A. Odoni (2009). Use of Queuing Models to Estimate Delay Savings from 4D Trajectory Precision. Eighth USA/Europe Air Traffic Management Research and Development Seminar (ATM 2009).

- [10] P. D. Charles (1977). The Output of a D/M/1 Queue, SIAM Journal on Applied Mathematics, Vol. 32, No. 3 (May, 1977), pp. 571-587.
- [11] P. Kivestu (1976). Alternative Methods of Investigating the Time-Dependent M/G/k Queue. Master's Thesis, MIT.